СЛУЧАЙНЫЕ ПОГРЕШНОСТИ И ЗАКОНЫ РАСПРЕДЕЛЕНИЯ

Лекція №2

Простейшим термометром успеха может служить зависть неудачниковПростейшим термометром успеха может служить зависть неудачников Простейшим термометром успеха может служить зависть неудачников Дали,

Случайные явления и вероятность события

- ***Случайное явление*** *- это такое явление, которое при неоднократном воспроизведении одного и того же опыта протекает каждый раз по разному. Факт, который в результате опыта может произойти или не произойти, называется событием. Для численного сравнения событий по степени их возможности вводится понятие вероятности события. В качестве единицы измерения вероятности принимают вероятность достоверного события, для которого вероятность равна единице. Пример достоверного события - доставание белого шара из урны, в которой лежат только белые шары.*
- *Событие, которое в данном опыте не может произойти, называется невозможным. Ему приписывается вероятность, равная нулю. Все другие вероятные события имеют вероятность больше нуля, но меньше единицы.*

Случайные явления и вероятность события

- Вероятность события Р вычисляется как отношение числа благоприятных случаев к общему числу случаев n
  - $P=m/n$
- Случай называется благоприятным некоторому событию, если появление этого случая влечет за собой появление данного события. Так, например, при бросании монеты вероятность появления герба равна 1/2.
- Если вероятность нельзя определить теоретически, то ее определяют статистически, т.е. находят частоту событий.
- Если проведена серия из n опытов и в m из них произошло некоторое событие, то частотой Р* называется отношение числа опытов, в которых появилось это событие, к общему числу опытов :
  - $P^*=m/n$

$\gamma = \frac{x - X}{X} = \frac{\Delta}{X}$

Случайные явления и вероятность события

- *При небольшом числе опытов частота носит случайный характер, но с увеличением числа опытов частота приближается к вероятности.*
- *На практике часто приходится иметь дело не с достоверными и невозможными событиями, а с практически достоверными и практически невозможными. Это события, вероятности которых близки соответственно к единице и нулю.*
- *Указанные события играют важную роль в теории вероятностей. На этих событиях строится принцип практической уверенности: если вероятность некоторого события в данном опыте весьма мала, то можно быть практически уверенным в том, что при однократном выполнении опыта это событие не произойдет. Этот принцип не может быть доказан математически, но подтверждается всем практическим опытом человека.*

Понятие о законе распределения случайной величины

- Сумма вероятностей всех возможных значений случайной величины (СВ) равна единице : $\sum_{i=1}^{n} P_i = 1$
- Докажем это на примерах:
- 1. При бросании монеты возможны два случая : выпадение герба или не герба. Вероятность каждого случая равна 1/2.
- 2. При бросании игральной кости вероятность появления любой из шести граней игральной кости равна 1/6, сумма всех вероятностей равна единице.
- Принято различать случайные величины прерывного (дискретно го) и непрерывного типа. Случайные величины, принимающие только отдельные друг от друга значения, которые можно заранее перечислить, называются прерывными или дискретными. Случайные величины , возможные значения которых непрерывно заполняют некоторый промежуток, называются непрерывными (например, длина отрезка, погрешность измерения).

Понятие о законе распределения случайной величины

- *Если X1,X2,...,Xn - возможные значения СВ X, а P1,P2,...,Pn - вероятность этих событий, то можно составить таблицу :*

X_i P_i	X_1 P_1	X_2 P_2	... ...	X_n P_n

- *Если сумма вероятностей данных значений равна единице, то с вероятностной точки зрения СВ полностью охарактеризована, т.е. установлен закон распределения СВ.*
- *Таким образом, закон распределения - всякое соотношение, устанавливающее связь между возможными значениями СВ и соответствующими им вероятностями. Приведенную таблицу называют рядом распределения.*

Понятие о законе распределения случайной величины

- *Для придания ряду наглядного вида по его данным строят многоугольник распределения.*
- *Непрерывные величины имеют бесчисленное множество возможных решений. Поэтому весь интервал возможных значений СВ разбивают на ряд меньших интервалов Ii и определяют частоту P* попадания в данный интервал. Получают статистический ряд.*

Ii	$X_1 - X_2$	$X_2 - X_3$	$X_3 - X_4$	....	$X_{n-1} - X_n$
P^*	P_1^*	P_2^*	P_3^*	....	P_n^*

Понятие о законе распределения случайной величины

- *По полученным данным строят прямоугольники, основанием которых служат выбранные интервалы, а высота (ордината) определяется как отношение частоты к длине этого интервала. Ступенчатая кривая, огибающая прямоугольники, называется гистограммой. Площадь под гистограммой равна единице. Если $\Delta x \to 0$, то ступенчатая кривая перейдет в плавную кривую - кривую распределения вероятности непрерывной СВ. Ординату любой точки кривой принято называть плотностью вероятности P(x).*
- *Для подсчета вероятности попадания СВ в заданный промежуток от a до в необходимо найти заштрихованную площадь как показано на рис.4.2.,б.*
- *Для примера рассмотрим два закона распределения - равномерный и нормальный (подробно о них смотри дальше).*

Понятие о законе распределения случайной величины

- Если непрерывная СВ X принимает значения в интервале от X1 до X2 с одной и той же плотностью вероятности (рис.4.3.), то такой закон называется равномерным и записывается так :

$$\begin{cases} P(x) = \dfrac{1}{x_2 - x_1} = const, & \text{при} \quad x_1 < x < x_2 \\ P(x) = 0, & \text{при} \quad x > x_1 \quad \text{и} \quad x > x_2 \end{cases}$$

- Для нормального закона распределения (закон Гаусса) (рис.4.4.) плотность вероятности описывается выражением

$$P(X) = \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{[x - M(x)]^2}{2\delta^2}}$$

- где M(x) - математическое ожидание;
- δ - среднеквадратическое отклонение (см.ниже).
- Для приведенного на рис.4.4. закона распределения M(x) = 0.

Понятие о законе распределения случайной величины

- Из формулы видно, что нормальный закон характеризуется двумя параметрами - математическим ожиданием и среднеквадратической погрешностью. Математическое ожидание определяет положение распределения по оси абсцисс и не влияет на форму кривой.
- Параметр δ является характеристикой рассеивания, от него зависит форма распределения. Наибольшая ордината, равная :

$$\frac{1}{\delta\sqrt{2\pi}}$$

- обратно пропорциональна δ. Поскольку площадь под кривой распределения всегда равна единице, то при увеличении δ кривая понижается и растягивается вдоль оси абсцисс.
- Из формы кривой распределения вытекают свойства СВ (погрешностей).

Понятие о законе распределения случайной величины

- *1. Наибольшая плотность вероятности соответствует погрешности, равной нулю.*
- *2. Погрешности, одинаковые по абсолютной величине, но с разными знаками, равновероятны (кривая симметрична).*
- *3. Чем больше погрешность, тем меньше вероятность ее появления. Среднеквадратическая погрешность соответствует такому значению δ, для которого 68% погрешностей будут меньше данной величины и 32% погрешностей будут больше данной величины δ.*
- *В теории погрешностей различают также средневероятную (вероятную) ρ и среднеарифметическую (по модулю) погрешность υ.*

Понятие о законе распределения случайной величины

- *Вероятная - это такая погрешность, которая с одинаковой вероятностью может быть как превзойдена, так и не достигнута. Иначе говоря, это такая погрешность, для которой 50% всех значений погрешности меньше ее, а другие 50% - больше.*
- *Среднеарифметическая (по модулю) погрешность соответствует такому значению погрешности, для которой 57.5% погрешности будут меньше υ и 42.5% - больше υ.*
- *Связаны между собой указанные погрешности следующими соотношениями : $\rho = 0.6745$; $\upsilon = 0.7979$*
- *Причина того, что многие СВ подчиняются нормальному закону распределения, заключается в следующем. Оказывается, что если СВ зависит от многих факторов, каждый из которых, взятый в отдельности, влияет на эту величину сравнительно мало, то суммарное влияние этих факторов приводит к нормальному распределению СВ.*

Числовые характеристики законов распределения

- *Для оценки свойств законов распределения используют числовые характеристики, называемые моментами. С помощью числовых характеристик можно решать вероятностные задачи, оставляя в стороне законы распределения.*
- *Основной характеристикой положения случайной величины на числовой оси, указывающей некоторое среднее значение, вокруг которого группируются все возможные значения случайной величины, является математическое ожидание M[X], называемое также первым начальным моментом.*
- *Для дискретных величин математическое ожидание определяется как сумма произведений всех возможных значений x_i на их вероятности P_i, т.е.*

$$M[X] = \sum_{i=1}^{m} x_i P_i$$

Числовые характеристики законов распределения

- а для непрерывных величин с плотностью распределения вероятности P(x) - как интеграл

$$M[X] = \int_{-\infty}^{+\infty} xP(x)dx$$

- Зная математическое ожидание, можно найти случайные отклонения Δi = xi - M[X] каждого результата наблюдения от математического ожидания.
- Начальным моментом n -го порядка дискретной величины X называется сумма вида

$$a_n[x] = M[X^n] = \sum_{i=1}^{i=m} x_i^{\ n} P_i$$

- а начальным моментом непрерывной случайной величины - интеграл вида

$$a_n[x] = M[X^n] = \int_{-\infty}^{+\infty} x^n P(x)dx$$

Числовые характеристики законов распределения

- Центральным моментом n-го порядка случайной величины называется математическое ожидание n-й степени соответствующей центрированной случайной величины:

$$\mu_n[X] = M[(x - M[X])^n]$$

- Второй центральный момент называется дисперсией случайной величины и характеризует рассеяние случайной величины вокруг математического ожидания:

$$\mu_2[X] = D[X] = M[(x - M[X])^2]$$

- Или иначе, для непрерывных величин

$$D[x] = \int_{-\infty}^{+\infty} (x - X)^2 P(x) dx ,$$

Числовые характеристики законов распределения

- для дискретных величин

$$D[x] = \sum_{i=1}^{n} (x_i - X)^2 P_i$$

- Дисперсия имеет размерность квадрата случайной величины и выражает как бы мощность ее рассеяния.
- Для наглядной характеристики рассеяния пользуются среднеквадратическим отклонением

$$\delta = \sqrt{D[X]}$$

- размерность которого совпадает с размерностью случайной величины. Величина δ характеризует действующее значение случайной величины.
- Третий центральный момент характеризует асимметрию, т. е. скошенность распределения. Для симметричных относительно математического ожидания законов распределения он равен нулю. Четвертый центральный момент характеризует форму, т.е. крутизну спадов распределения, а его относительное значение $\varepsilon = \frac{\mu_4}{\delta_4}$

Числовые характеристики законов распределения

- *называется эксцессом (изменяется от 1 до ∞)*
- *Часто пользуются величиной*

$$S = \frac{1}{\sqrt{\varepsilon}}$$

- *или*

$$S = \frac{D}{\sqrt{\mu_4}} = \sqrt{\frac{\delta_4}{\mu_4}}$$

- *называемой контрэксцессом (изменяется от 0 до 1).*

Экспериментальное определение числовых характеристик случайных величин

- *При проведении эксперимента получают не полный набор всех значений случайной величины, которой в статистике называют генеральной совокупностью, а лишь некоторое число этих значений, которое называется выборкой.*
- *Определенные по этим данным характеристики закона распределения являются приближенными, их называют оценками. Особенность экспериментальных значений случайной величины состоит в том, что они являются естественно квантованными (округленными).*
- *Эмпирическое значение вероятности попадания значений случайной величины в i-тый интервал группирования определяется, как указывалось ранее, через частоту (или частность) P*(x).*

$$P_i(x) = \frac{P_i^*(x)}{d} = \frac{n_i}{nd}$$

Экспериментальное определение числовых характеристик случайных величин

- где ni - количество xi, попавших в i-тый интервал;
- n - полное число наблюдений;
- d - длина интервала.
- Оценка математического ожидания случайной величины есть среднее арифметическое всех полученных результатов наблюдений, т.е.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Несмотря на конечный объем выборки, эта оценка не имеет систематической составляющей погрешности, т.е. является несмещенной оценкой.
- Оценка дисперсии D* производится по формуле

$$D^* = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2$$

Экспериментальное определение числовых характеристик случайных величин

- Она оказывается смещенной, т.е. кроме разброса имеет систематическую отрицательную погрешность, возрастающую по мере уменьшения n.
- Для исключения этой погрешности найденное значение D* умножают на поправочный множитель Бесселя.

$$A = \frac{n}{n-1}$$

- Таким образом

$$\overline{D}^*(x_i) = \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n-1}$$

- Соответственно среднеквадратическое отклонение

$$\overline{\delta}^*(x_i) = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n-1}}$$

- Оценка четвертого центрального момента (без поправки на смещение)

$$\overline{\mu_4}^* = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^4$$

Максимальная или предельная оценка случайной погрешности

- *Для оценки максимальной или предельной погрешности обычно берут наибольшее по модулю значение погрешности, встретившееся в данном произвольно ограниченном ряде наблюдений. Основное преимущество этой оценки состоит в простоте ее определения - из всех зарегистрированных отклонений выбирается наибольшее (без учета знака) и принимается в качестве "предельного" значения. Однако такая "предельная" оценка имеет существенные недостатки.*
- *Во-первых, она случайна. Если в данной серии наблюдений ока залась максимальной погрешность, например, 5%, то это не гарантирует того, что в следующей серии она не окажется раной 6 или 4%.*
- *Во-вторых, такая оценка существенно зависит от объема выборки, т.е. серии наблюдений, поскольку при продолжении наблюдений всегда могут встретиться еще большие погрешности.*

Максимальная или предельная оценка случайной погрешности

- *Например, при нормальном законе распределения погрешности $\Delta m = \delta$ встречаются в среднем один раз на каждые 3 наблюдения, $\Delta m = 2\delta$ -один раз на 22 наблюдения, $\Delta m = 3\delta$ - в среднем один раз на 370 наблюдений и $\Delta m = 4\delta$ - один раз на 15000 наблюдений.*
- *Это создает неодинаковые условия аттестации простых и сложных измерительных устройств, которые проверяются при различном числе наблюдений (например, десятки и сотни).*
- *И, наконец, при оценке результирующей погрешности средств измерений, состоящих из нескольких отдельных устройств, совершенно бессмысленно суммировать "предельные" значения статистически независимых составляющих погрешности.*

Максимальная или предельная оценка случайной погрешности

- *Покажем это на примере. Пусть средство измерений имеет три составляющие погрешности, каждая из которых может принимать 11 различных значений: -5, -4, ..., +4, +5.*
- *Тогда общее число возможных комбинаций 113 = 1331. Суммирование предельных значений соответствует только двум случаям, когда все три погрешности равны +5 или -5.*
- *Если принять, что все комбинации равновероятны (на самом деле большие погрешности встречаются реже), то вероятность этих двух случаев равна 2/1331 = 1/165.*
- *Если же суммируются не три, а десять таких же погрешностей, то вероятность результирующей погрешности, равной сумме их предельных значений, составит 10-10, т.е. подобная ситуация практически никогда не встретится.*

Максимальная или предельная оценка случайной погрешности

- *Такая "перестраховочная" оценка для наилучшего стечения обстоятельств относится к несуществующей ситуации. Поэтому результирующую погрешность никогда не определяют путем суммирования "предельных" погрешностей, а находят экспериментально, либо расчетным путем, но используют при этом другие характеристики.*

Доверительная погрешность и доверительная вероятность

- *Как было показано ранее, площадь под кривой плотности распределения равна единице, т.к. отражает вероятность всех возможных погрешностей. Эту площадь можно разделить на некоторые части вертикальными линиями. Абсциссы таких линий называются квантилями. Квантиль β%- это такая абсцисса, площадь слева от которой составляет β% общей площади под кривой P(x). Медиана - это 50%-ная квантиль, которая делит площадь на две равные части. Между 5%-ной и 95%-ной квантилями заключено 90% всех погрешностей.*
- *Разность двух квантилей называется интерквантильным промежутком, а половину интерквантильного промежутка обычно принимают за доверительное значение погрешности*

$$\Delta_\partial = \frac{d_\partial}{2}$$

Доверительная погрешность и доверительная вероятность

- *Т.к. доверительный интервал выбирается произвольно, то при указании доверительного значения погрешности необходимо указывать одновременно и значение так называемой доверительной вероятности Рд, т.е. вероятности того, что модуль погрешности будет не больше значения Δд. Так, для нормального закона распределений при Рд = 0.9545 доверительная погрешность соответствует Δд = 2δ, а при Рд = 0.997. Δд = 3δ.*
- *Достоинство доверительной погрешности состоит в том, что ее значение может быть довольно просто определено по экспериментальным данным. Однако, чем с большей вероятностью Рд мы хотим определить Δд, тем больше экспиревентов нам следует провести*
-

$P_д$	0.8	0.9	0.95	0.99	0.995	0.997
n	10	20	40	200	400	667

Доверительная погрешность и доверительная вероятность

- *Из таблицы видно, что практически можно определить значения Δд лишь с доверительной вероятностью Рд ≤ 0.95. Иначе говоря, если не известен закон распределения погрешностей, то при малом числе испытаний (20-30) какие - либо сведения о ходе кривой в районе с Рд > 0.95 отсутствуют. Тем не менее иногда допускают такую ошибку. На основании 20-30 наблюдений вычисляют среднеквадратическое отклонение δ , делают предположение о нормальности закона распределения и указывают на предельную погрешность Δд = Δт = 3δ с доверительной вероятностью Рд = 0.997. Как будет показано далее, реаль ные законы распределения погрешностей очень часто далеки от нормального.*
- *Основным недостатком доверительной погрешности Δд , как и "максимальной" Δт , является невозможность их суммирования.*

Образование композиций законов распределения

- *Результирующие погрешности средств измерений складываются из ряда составляющих, а при сложении погрешностей законы их распределения существенно деформируются. Закон распределения P(x) = P(X1 + X2) суммы двух независимых случайных величин, имеющих распределения P1(x) и P2(x), называется композицией законов распределения.*
- *Покажем это на примерах. При суммировании двух равномерно распределенных случайных величин X1 и X2 образование композиций можно представить как размыв резко ограниченных концов широкого распределения (шириной а) на величину протяженности в менее широкого распределения. Полученный закон распределения имеет форму трапеции. Аналогично получается композиция равномерного и нормального распределений.*

Образование композиций законов распределения

- *Композиция двух одинаковых равномерных распределений является треугольной (распределение Симпсона), т.к. в этом случае верхнее основание трапеции обращается в нуль, а нижнее - в 2а.*
- *Композиция двух треугольных распределений описывается участками парабол и имеет также удвоенную ширину основания. Приведенные рисунки построены без соблюдения относительного масштаба кривых по вертикали. Этот масштаб должен быть таким, чтобы в каждом случае площадь под кривой закона распределения была равна единице*

Значение среднеквадратической погрешности для оценки суммарной погрешности

- Как будет показано далее, реальные законы распределения пог решности приборов весьма разнообразны и часто далеки от нормального. Для оценки суммарной погрешности прибора не могут быть использованы ни предельная Δm , ни доверительная погрешность $\Delta д$. Для этого используют среднеквадратическую погрешность.
- Среднеквадратическое значение δ случайной величины - это ее действующее, эффективное значение, подобное эффективному (в энергетическом смысле) значению тока *i(t)* со сложной формой кривой

$$I = \sqrt{\frac{P}{R}} = \sqrt{\frac{1}{T}\int_0^T i^2(t)dt}$$

- Аналогично этому

$$\delta = \sqrt{D} = \sqrt{\frac{1}{T}\int_{-\infty}^{+\infty} x^2 P(x)dt}$$

- *где D - дисперсия; P(X) - плотность распределения.*

Значение среднеквадратической погрешности для оценки суммарной погрешности

- *Достоинством такой энергетической оценки случайной величины с произвольным законом распределения является возможность оценки суммарного действия нескольких таких независимых случайных величин благодаря свойству скалярного суммирования мощностей. Так, если по участку цепи протекает несколько статистически независимых токов с любой формой кривых, то суммарная их мощностьпросто равна сумме токов мощностей PE = P1 + P2 + ... + Pn , а действующее значение суммы токов определяется из соотношения I2E = I21 + I22 +...+ I2n . При этом токи Ii должны быть представлены их действующими (а не максимальными) значениями. Подобно этому действию значение суммы статистически независимых величин равно*

$$D_E = \sum_{i=1}^{n} D_i^{\,2} \qquad \text{или} \qquad \delta_E^{\,2} = \sum_{i=1}^{n} \delta_i^{\,2}$$

Значение среднеквадратической погрешности для оценки суммарной погрешности

- *независимо от законов распределения суммируемых величин. Таким образом, для суммирования отдельных составляющих погрешностей средств измерений они (погрешности) должны быть предварительно представлены среднеквадратическими значениями. При этом представляется возможность не только производить расчет (анализ) суммарной погрешности. но и вычитать погрешности, т.е. производить синтез, что необходимо при разработке устройств с заданной погрешностью.*
- *Действительно, если* $\delta_E = \sqrt{\delta_1{}^2 + \delta_2{}^2}$ то $\delta_2 = \sqrt{\delta_E{}^2 - \delta_1{}^2}$
- *Правда, при этом нужно учитывать корреляционные взаимосвязи суммируемых погрешностей. Полученные здесь выражения для среднеквадратической погрешности ряда (т.е. серии) случайных величин и погрешность математи-ческого ожидания (т.е. среднего значения) этого ряда случайных величин связаны между собой отношением* $S = \dfrac{\delta}{\sqrt{n}}$

Значение среднеквадратической погрешности для оценки суммарной погрешности

- *Соответственно средняя (вероятная) погрешность математического ожидания равна*

$$R = \frac{\rho}{\sqrt{n}}$$

- *и средняя по модулю арифметическая погрешность математического ожидания составит величину*

$$Q = \frac{\upsilon}{\sqrt{n}}$$

Спасибо за внимание!